Faculty of Computer Science and Information Technology

Multi-Agent Approach for Occlusion Handling in Crowded Surveillance Video

Lee Beng Yong

Doctor of Philosophy
2018

Multi-Agent Approach for Occlusion Handling In Crowded Surveillance Video

Lee Beng Yong

A thesis submitted

In fulfilment of the requirements for the degree of Doctor of Philosophy

(Computer Science)

Faculty of Computer Science and Information Technology
UNIVERSITI MALAYSIA SARAWAK
2018

UNIVERSITI MALAYSIA SARAWAK

Grade: _____

Please tick (√)

Final Year Project Report ☐

Masters ☐

PhD ☑

DECLARATION OF ORIGINAL WORK

This declaration is made on the ..2.. day of ..Aug.. ~~2016.~~ 2018

Student's Declaration:

I LEE BENG YONG (10011481) of Faculty of Computer Science and Information Technology

(PLEASE INDICATE STUDENT'S NAME, MATRIC NO. AND FACULTY) hereby declare that the work entitled, Multi-Agent Approach for Occlusion Handling in Crowded Surveillance Video is my original work. I have not copied from any other students' work or from any other sources except where due reference or acknowledgement is made explicitly in the text, nor has any part been written for me by another person.

2 Aug 2018
Date submitted

Lee Beng Yong (10011481)
Name of the student (Matric No.)

Supervisor's Declaration:

I Prof. Dr Wang Yin Chai (SUPERVISOR'S NAME) hereby certifies that the work entitled, Multi-Agent Approach for Occlusion in Crowded Surveillance Video (TITLE) was prepared by the above named student, and was submitted to the "FACULTY" as a * partial/full fulfillment for the conferment of Ph.D (PLEASE INDICATE THE DEGREE), and the aforementioned work, to the best of my knowledge, is the said student's work

Received for examination by: Wang Yin Chai
(Name of the supervisor)

Date: 2 Aug 2018

I declare this Project/Thesis is classified as (Please tick (√)):

☐ **CONFIDENTIAL** (Contains confidential information under the Official Secret Act 1972)*
☐ **RESTRICTED** (Contains restricted information as specified by the organisation where research was done)*
☑ **OPEN ACCESS**

Validation of Project/Thesis

I therefore duly affirmed with free consent and willingness declared that this said Project/Thesis shall be placed officially in the Centre for Academic Information Services with the abide interest and rights as follows:

- This Project/Thesis is the sole legal property of Universiti Malaysia Sarawak (UNIMAS).
- The Centre for Academic Information Services has the lawful right to make copies for the purpose of academic and research only and not for other purpose.
- The Centre for Academic Information Services has the lawful right to digitise the content to for the Local Content Database.
- The Centre for Academic Information Services has the lawful right to make copies of the Project/Thesis for academic exchange between Higher Learning Institute.
- No dispute or any claim shall arise from the student itself neither third party on this Project/Thesis once it becomes sole property of UNIMAS.
- This Project/Thesis or any material, data and information related to it shall not be distributed, published or disclosed to any party by the student except with UNIMAS permission.

Student's signature _____ 2 Aug 2018
(Date)

Current Address: Lot 2089 Lorong 1
Jalan Kong Phin, 93250, Kuching.

Supervisor's signature: _____ 2/8/18
(Date)

PROF. DR. YIN CHAI
Director
Centre of Excellence for Image Analysis
And Spatial Technologies
Universiti Malaysia Sarawak

Notes: * If the Project/Thesis is CONFIDENTIAL or RESTRICTED, please attach together as annexure a letter from the organisation with the period and reasons of confidentiality and restriction.

[The instrument was duly prepared by The Centre for Academic Information Services]

DECLARATION

I hereby declare that the thesis is based on my original work except for quotations and citation, which have been duly acknowledged. The thesis has not been accepted for any degree and is not concurrently submitted in the candidature of any other degree.

_____

Name of the student: Lee Beng Yong

Date: 7 May 2018

DEDICATION

To my wife, Liew, Leehung, my daughter, Joanne Lee and my parents for their unconditional love and support.

ACKNOWLEDGMENT

I would like to express my sincere gratitude to my supervisor Professor Dr Wang Yin Chai and my co-supervisors Dr Cheah Wai Siang and Associate Professor Dr Noor Alamshah Bolhassan. It has been a great pleasure to work with them and to learn from their experience. Their patience and support guided me through my PhD study. Also, I would like to thank my colleagues and friends in the Faculty of Computer Science and Information Technology with whom I had the pleasure to work and discuss.

ABSTRACT

The task of object tracking is the key to the effective use of more advanced technologies such as event recognition in the surveillance video. To recognise an event, people and objects must be tracked. Tracking also enhances the performance of tasks such as crowd analysis or human identification. Among all existing object tracking methods, Continuously Adaptive Mean-shift (CAMShift) is one of the most popular methods adopted by many researchers. However, to use CAMShift to track an object, one should also verify that the chosen target representation is a sufficient discriminator for the application domain to ensure CAMShift can efficiently work in its application. This thesis investigates improvements to the performance of object tracking algorithms through improved target representation selection using a novel multi competitive agents system that consists of a manager agent, an evaluator agent and multiple CAMShift tracker agents. The interaction between the manager agent and the CAMShift tracker agents in the proposed is performed via the Call-For-Proposal protocol. Through the interaction, the CAMShift tracker agents in the system can compete among themselves to select an optimum target representation. The proposed algorithm is tested on lab simulation videos and shows significant improvement in tracking. Kalman filter is then added to the CAMShift tracker agents to solve full occlusion problem. When a tracked object is fully occluded, it is not visible in the video scenes. CAMShift tracker agent will not be able to retrieve any fully occluded object representation model from the scene, and therefore Kalman filter is used to predict the location of the tracked object based on spatial motion characteristic of the tracked object in previous frames. The proposed system is then tested on two famous benchmark surveillance video dataset, namely the PETS 2009 and CAVIAR. The results

show that the proposed system can track objects in crowded scenes even when a full occlusion occurs during the tracking process.

Keywords: Continuously Adaptive Mean-shift, Kalman Filter, Multi-Agents System Occlusion, Crowded Surveillance Video, Object Tracking

Kaedah Multi Ejen untuk Pengendalian Halangan bagi Penjejakan Objek dalam Video Pengawasan yang Sesak

ABSTRAK

Tugas penjejakan objek adalah kunci kepada keberkesanan penggunaan teknologi yang lebih maju seperti pengecaman peristiwa di video pengawasan. Untuk pengecaman peristiwa, manusia dan objek perlu dikesan. Penjejakan objek juga dapat meningkatkan prestasi tugas-tugas seperti menganalisa kelakuan manusia atau pengenalan manusia. Di antara semua kaedah penjejakan objek yang sedia ada, kaedah Peralihan Purata Berpenyesuaian Berterusan (CAMShift) adalah salah satu kaedah yang paling popular digunakan oleh kebanyakan penyelidik. Walau bagaimanapun, untuk menggunakan CAMShift untuk menjejak objek, seseorang itu perlu juga mengesahkan bahawa perwakilan sasaran yang dipilih adalah cukup diskriminan untuk domain aplikasi bagi memastikan CAMShift boleh bekerja dengan berkesan dalam aplikasi mereka. Tesis ini mengkaji cara penambahbaikan untuk prestasi algoritma penjejakan objek melalui pertambahbaikan dalam pemilihan perwakilan sasaran. Satu sistem multi ejen berkompetitif novel yang terdiri daripada ejen pengurus, ejen penilai dan ejen-ejen pengesan CAMShift yang mampu untuk bersaing dengan satu sama lain dalam usaha untuk pilih sasaran perwakilan optimum telah dicadangkan. Interaksi antara ejen-ejen dalam sistem yang dicadangkan adalah dilaksanakan melalui protokol Memanggil Untuk Cadangan (CFP). Algoritma yang dicadangkan ini telah diuji dengan menggunakan video simulasi makmal dan ianya menunjukkan peningkatan prestasi pengesanan yang ketara berbanding dengan sistem asas. Kemudian, penapis Kalman telah ditambah ke dalam ejen CAMShift tracker untuk menyelesaikan masalah halangan penuh. Apabila objek yang

dikesan adalah dihalang sepenuhnya, ia tidak boleh dilihat dalam babak video. ejen CAMShift tracker tidak akan dapat apa-apa model perwakilan objek dari babak video dan oleh itu penapis Kalman telah digunakan untuk meramalkan lokasi objek dikesan berdasarkan ciri-ciri gerakan spatial objek dikesan dalam bingkai sebelumnya. Sistem yang dicadangkan ini kemudian diuji dengan menggunakan dua set data pengawasan video penanda aras yang terkemuka, iaitu PETS 2009 dan CAVIAR. Keputusan menunjukkan bahawa sistem yang dicadangkan boleh mengesan objek dalam adegan sesak walaupun halangan penuh berlaku semasa proses pengesanan.

Kata kunci: *Peralihan Purata Berpenyesuaian Berterusan, Penapis Kalman, Sistem Multi Agen, Halangan, Video Pengawasan yang Sesak, Penjejakan Objek*

TABLE OF CONTENTS

Page

DECLARATION i

DEDICATION ii

ACKNOWLEDGMENT iii

ABSTRACT iv

ABSTRAK vi

TABLE OF CONTENTS viii

LIST OF TABLES xii

LIST OF FIGURES xiv

CHAPTER 1: INTRODUCTION 1

1.1 Introduction 1

1.2 Background of Study 2

1.3 Research Questions 5

1.4 Aims and Objectives 5

1.5 Research Scope 6

1.6 Outlines of Thesis 6

CHAPTER 2: LITERATURE REVIEW 9

2.1 Introduction 9

2.2 Object Tracking 9

2.3 Kernel-Based Object Tracking 12

2.3.1 Continuously Adaptive Mean-Shift (CAMShift) Tracking 13

2.3.2 CAMShift Limitation 19

2.4 Occlusion 19

2.5 Occlusion Handling 22

2.5.1 Depth Analysis 23

2.5.2 Optimal Camera Placement Method 24

2.5.3 Motion Prediction 25

2.5.4 Occlusion Handling Limitation 31

2.6 Multi Models Object Tracking 32

2.7 Multi-Agent System in Object Tracking 33

2.7.1 Previous Work of MAS in Object Tracking System 35

2.7.2 Role-Oriented Analysis and Design for Multi-Agent Programming (ROADMAP) 40

2.7.3 Negotiation Mechanism in Object Tracking 41

2.8 Surveillance Video Benchmark Dataset 43

2.9 Measurement of Detection Accuracy 48

2.10 Summary 51

CHAPTER 3: MULTI CAMSHIFT AGENT FRAMEWORK FOR OCCLUSION HANDLING IN OBJECT TRACKING 54

3.1 Introduction 54

3.2 Goal Model 55

3.3 Role Model 56

3.4 Organisation Model 59

3.5 Domain Model 59

3.6 Agent-Role Coupling 61

3.7 Interaction Model 63

3.8 Knowledge Model 67

3.9 Behaviour Model 68

3.10 Summary 74

CHAPTER 4: IMPLEMENTATION 75

4.1 Introduction 75

4.2 Negotiation Mechanism in Object Tracking 76

4.3 Implementation of JADE Based Tracking Manager Agent 76

4.4 Implementation of JADE Based Object Tracker Agent 83

4.5 Implementation of JADE Based Proposal Evaluator Agent 87

4.6 Simulation of Object Tracking System in JADE Platform 89

4.7 Summary 93

CHAPTER 5: RESULTS AND DISCUSSION 94

5.1 Introduction 94

5.2 Preliminary Experiment 94

5.2.1 Simulation Video Data 95

5.2.2 Tracking Performance Measurement 98
5.2.3 Evaluation Methods 98
5.2.4 Preliminary Experiment Results 99
5.2.5 Effect of Occlusion 102
5.2.6 Finding of Preliminary Experiment 105
5.3 Benchmark Dataset 105
5.4 Results and Analysis 106
5.4.1 Qualitative Result 107
5.4.2 Quantitative Result 111
5.5 Agents Communication in JADE 116
5.6 Summary 120
CHAPTER 6: CONCLUSION AND LIMITATION 122
6.1 Contributions 122
6.2 Future Work 124
REFERENCES 125
APPENDICES 145

LIST OF TABLES

Page

Table 2.1 Reviewed papers that use MAS for object tracking. 39

Table 2.2 Common performative used in a negotiation mechanism 42

Table 2.3 Reviewed benchmark dataset 43

Table 3.1 Role model for user 57

Table 3.2 Role model for object tracking manager 57

Table 3.3 Role model for object tracker 58

Table 3.4 Role model for proposal evaluator 58

Table 3.5 Descriptor of the type Tracking manager agent 62

Table 3.6 Descriptor of the type Object tracker agent 62

Table 3.7 Descriptor of the type Proposal evaluator agent 62

Table 3.8 A scenario for managing object tracking 69

Table 3.9 A scenario for tracking object 71

Table 3.10 The models of R1 until R6 for tracking object 72

Table 5.1 Video label and description 96

Table 5.2 Tracking result (SFDA) for six simulation videos 99

Table 5.3 Tracking result (SFDA) for the PETS 2009 and CAVIAR Video 111

Table 5.4 The lower boundary (_vmin) and upper boundary (_vmax) assigned for each object tracking agents when a different number of agents is used 112

Table 5.5 Tracking result on simulation video A1 dataset by using a different number of “CAMShiftAgents” 113

Table 5.6 Tracking result on simulation video A3 dataset by using a different number of “CAMShiftAgents” 113

Table 5.7 Tracking result on PETS 2009 view 001 S1 datasets by using a different number of “CAMShiftAgents” 115

Table 5.8 Tracking result on CAVIAR atrium dataset 116

LIST OF FIGURES

Page

Figure 1.1 The Essential process of Intelligent video surveillance System 3

Figure 2.1 Object representations. (a) Centroid, (b) multiple points, (c) rectangular patch, (d) elliptical patch, (e) part-based multiple patches, (f) object skeleton, (g) complete object contour, (h) control points on object contour, (i) object silhouette (Yilmaz, Javed, & Shah, 2006). 13

Figure 2.2 An ellipsoidal region represents a human target in the PETS 2009 video frame..... 14

Figure 2.3 Steps for generating normalised colour histogram within the target (track window)..... 15

Figure 2.4 Example of the colour histogram of a target. 16

Figure 2.5 Back Projection of the target object in the video frame. 17

Figure 2.6 Mean-shift tracker window always has the same size when target car is farther away, and it is very close to the camera (OpenCV, 2017). 18

Figure 2.7 The state vector of a moving robot is described as its position (x, y) and its direction heading (θ). Source (Baltzakis, 2004)..... 26

Figure 2.8 The initial estimate values for $\hat{x}$ and P. (Source: Ulusoy, 2017)..... 28

Figure 2.9 The initial estimate values for $\hat{x}$ and P. (Source: Ulusoy, 2017)..... 28

Figure 2.10 A frame from CAVIAR entrance lobby video 44

Figure 2.11 Sample frames from CAVIAR shopping mall videos..... 44

Figure 2.12 Sample images from PETS 2009 video dataset. 46

Figure 2.13 Ground truth data of PETS 2009 S1L1 view 001. 47

Figure 3.1 Goal model for object tracking 56
Figure 3.2 Organisation model for object tracking 59
Figure 3.3 The domain model for object tracking 60
Figure 3.4 The agent-role coupling diagram for object tracking 61
Figure 3.5 Interaction-frame diagram for object tracking 65
Figure 3.6 Interaction-sequence diagram for object tracking 66
Figure 3.7 Knowledge model for object tracking multi-agents system 68
Figure 3.8 Behaviour model for object tracking multi-agent system 73
Figure 4.1 “VideoAgent” graphical user interface 77
Figure 4.2 Extracted “VideoAgent” implementation code to let the user choose a targeted object 79
Figure 4.3 Extracted “VideoAgent” implementation code to search for object tracker agents from directory facilitator (df) agent 81
Figure 4.4 Extracted “VideoAgent” implementation code calling for a proposal to track targeted object 82
Figure 4.5 Extracted “VideoAgent” implementation code to determine the best-tracking result 83
Figure 4.6 Extracted “CamshiftAgent” implementation code for registration and resources setup 84
Figure 4.7 Extracted “CamshiftAgent” implementation code for responding the CFP received 87
Figure 4.8 Extracted “EvaluatorAgent” implementation code for registration 88
Figure 4.9 Extracted “Evaluator Agent” implementation code for replying back the evaluation score 89

Figure 4.10 JADE Agent Management GUI 90
Figure 4.11 Tracking Results (Frame No. 1, 20, 40, 60, 80, 100: Order from left to right, top to bottom) based on parameter (_vmin = 1, v_max = 128) 91
Figure 4.12 Tracking Results (Frame No. 1, 20, 40, 60, 80, 100: Order from left to right, top to bottom) based on parameter (_vmin = 129, v_max = 256) 92
Figure 5.1 Video Frame (a) Frame no. 1, (b) Frame no. 10, (c) frame no. 22, (d) frame no. 30 and (e) frame no. 50 of simulation video A1. 97
Figure 5.2 Video Frame (a) Frame no. 1, (b) Frame no. 10, (c) frame no. 22, (d) frame no. 30 and (e) frame no. 50 of simulation video A2. 97
Figure 5.3 Video Frame (a) Frame no. 1, (b) Frame no. 10, (c) frame no. 12, (d) frame no. 13, (e) frame no. 22, (f) frame no. 30 and (g) frame no. 50 of simulation video A3. 97
Figure 5.4 Video Frame (a) Frame no. 1, (b) Frame no. 10, (c) frame no. 15, (d) frame no. 22, (e) frame no. 30 and (f) frame no. 50 of simulation video A5. 98
Figure 5.5 Particle distribution: a) particles cover a large area at the initial state; b) when object trajectory remains consistent between frames, the particle area shrunk; c) particle area become so small and fail to detect the moving object change direction..... 100
Figure 5.6 Frame Detection Accuracy of Frame No. 12 to Frame no. 31 for Video Sequence A3..... 103
Figure 5.7 Frame Detection Accuracy of Frame No. 12 to Frame no. 31 for Video Sequence A4..... 103
Figure 5.8 Frame Detection Accuracy of Frame No. 12 to Frame no. 31 for Video Sequence A5..... 104

Figure 5.9 Frame Detection Accuracy of Frame No. 12 to Frame no. 31 for Video Sequence A6 105

Figure 5.10 Track result from CAMShift Tracker in MATLAB (Frame number 15, 20, 25, 30 (Left to right). 107

Figure 5.11 Track result from CAMShift Tracker agent (Frame number 15, 20, 25, 30 (Left to right). 108

Figure 5.12 Track result from CAMShift Tracker in MATLAB (Frame number 1, 50, 70, 90, 110 and 130. Left to right, top to bottom). 109

Figure 5.13 Track result from CAMShift Tracker agent (Frame number 1, 50, 70, 90, 110 and 130. Left to right, top to bottom) with parameter (_vmin = 25 and _vmax = 50). 110

Figure 5.14 Interaction between *CAMShift* agents, *df* agents and *video01* agent. 118

Figure 5.15 Message from Video agent to Directory Facilitator agent requesting for agent Id that provides object tracking service. 118

Figure 5.16 Content of *Request* message from *video01* agent to *df* agent. 119

Figure 5.17 Content of *Inform* message from *df* agent to *video01* agent. 119

CHAPTER 1
INTRODUCTION

1.1 Introduction

Due to the rapid drop in the cost of video cameras in recent years and the rise in security awareness, demand on surveillance has shown a steady increase from time to time (Cropley, 2016). In general, conventional video surveillance networks are typically monitored by one or more security personnel, looking at several monitors showing footage recorded from several cameras. Therefore, it is very likely to have missed important events as they happen because it is tough for a human to monitor all these cameras at the same time (Ali & Dailey, 2012).

Recently, many research on intelligent video surveillance have begun to address ways to automatically process some of the data acquired from surveillance video, to assist human operators to effectively detect events (Mathur & Bundele, 2016) such as people fighting in a building (Fisher, 2005), loitering, theft, and unattended luggage in subway (Computational Vision Group, Reading University, 2007) as they happen. The success of event recognition in intelligent video surveillance relies heavily on the fundamental processes in the system such as motion detection and objects tracking in the video surveillance (Mathur & Bundele, 2016).

As a result, many types of research on object tracking have been conducted recently to promote efficiency and effectiveness in object tracking especially in a difficult situation such as when target tracked object is occluded (Yilmaz, Javed, & Shah, 2006). This research attempts to improve the process of object tracking through design and implement

a parameter selection method for Multi Continuously Adaptive Mean-shift (CAMShift) object tracking algorithm.

In this chapter, the background of the intelligent video surveillance is presented in the next section before proceeding to a discussion of the research question and research objectives. The background study provided a brief explanation of what are the essential processes that involve in an intelligent video surveillance system. The background information on intelligent video surveillance help to identify the underlying problems in current work and lead to the objectives of this research.

The objectives of this research are listed in this chapter too. The objectives describe the motivations of this research. Next, the scope of the research is discussed in the last section of this chapter where the outlines of the research are listed.

1.2 Background of Study

Intelligent video surveillance involves the study of computer vision techniques to detect, track and understand objects and events in video monitoring (Pérez, García, Berlanga, & Molina, 2005; Mathur & Bundele, 2016). A common intelligent video surveillance consists of several important processes, namely the video acquisition, video recording, video compression, object detection, object tracking and event recognition as shown in Figure 1.1.

Video acquisition in intelligent video surveillance relies on hardware technology for capturing video images of higher resolutions while recording and compression are the arts of creating more efficient methods to store and index more video information into smaller storage and faster retrieval. Object detection with event recognition as the highest-level

processing in intelligent video surveillance, that is, the effectiveness of event recognition relies heavily on the accuracy of object tracking.

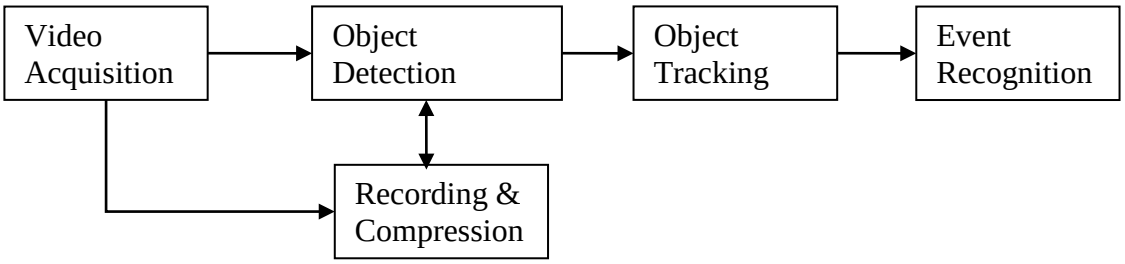

Figure 1.1: The Essential process of Intelligent video surveillance System

For instance, to recognise the path a person has walked through in a building, the system must first track the person based on the person's appearance from entry to exit. Another example is to recognise abnormal activity such as a person running in a subway station and a vehicle driven in the opposite direction; all these activities require accurate object tracking.

Object tracking is the process of following one or more objects in a video from the object's first appearance until the object exit the video scene. Some applications of object tracking system require the tracking system to continue the tracking if the object reappears in the video scene within a predefined time elapse. In other words, tracking can be defined as the problem of estimating the trajectory of an object in the image plane as it moves around a scene. A tracker assigns consistent labels to the tracked objects in different frames of a video. Additionally, depending on the tracking domain, a tracker can also provide object-centric information, such as orientation, area, or shape of an object. The tracked object in object tracking can be anything of interest within the scene, from human to a soccer ball and the vehicle.